DOCUMENT RESUME

ED 395 989 TM 025 183

AUTHOR Yu, Chong Ho; And Others
TITLE Identification of Misconceptions in the Central Limit Theorem and Related Concepts and Evaluation of Computer Media as a Remedial Tool.
PUB DATE 19 Apr 95
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Simulation; *Computer Software; Hypothesis Testing; *Identification; *Sampling; Statistical Analysis; *Statistical Distributions
IDENTIFIERS *Central Limit Theorem; *Misinterpretation

ABSTRACT

Central limit theorem (CLT) is considered an important topic in statistics, because it serves as the basis for subsequent learning in other crucial concepts such as hypothesis testing and power analysis. There is an increasing popularity in using dynamic computer software for illustrating CLT. Graphical displays do not necessarily clear up misconceptions related to this theorem. Many interactive computer simulations allow users to explore the programs in a "what-if" manner. However, users may further build up other misconceptions when they start with unclear concepts of the components that contribute to CLT. This paper analyzes common misconceptions in each component of CLT and evaluates the appropriateness of use of computer simulation. CLT states that a sampling distribution, which is the distribution of the means of random samples drawn from a population, becomes closer to normality as the sample size increases, regardless of the shape of the distribution. Misconceptions are found about the following areas: (1) randomness and random sampling; (2) relationships among sample, population, and sampling distribution; (3) normality; (4) parameters of the sampling distribution; and (5) relationships between the sampling distribution and hypothesis testing. (Contains 31 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 395 989

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

CHONG HO YU

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Identification of Misconceptions in the Central Limit Theorem and Related Conceptsand Evaluation of Computer Media as a Remedial Tool

Dr. Yu, Chong Ho and Dr. John T. Behrens
Arizona State University

Spencer Anthony
University of Oklahoma

Paper presented at the Annual Meeting of
the American Educational Research Association

April 19, 1995

Dr. Yu, Chong Ho
College of Education
159 Payne Hall
Arizona State University
Tempe AZ 85287-0611
Voice: (602)965-5039
Fax: (602)965-7058
email: alex.yu@asu.edu

TM025183

RUNNING HEAD: CLT

BEST COPY AVAILABLE

Abstract

Central limit theorem (CLT) is considered an important topic in statistics, because it serves as the basis for subsequent learning in other crucial concepts such as hypothesis testing and power analysis. There is an increasingly popularity in using dynamic computer software for illustrating CLT. Graphical displays do not necessarily clear up misconceptions related to this theorem. Many interactive computer simulations allow users to explore the programs in a "what-if" manner. However, users may further build up other misconceptions when they start with unclear concepts of the components that contribute to CLT. In this paper, we analyze the common misconceptions in each component of CLT, and evaluate the appropriateness of use of computer simulation.

Identification of Misconceptions in the Central Limit Theorem and Related Concepts and Evaluation of Computer Media as a Remedial Tool

Central limit theorem (CLT) is considered an important topic in statistics, because it serves as the basis for subsequent learning in other crucial concepts such as hypothesis testing and power analysis. CLT is said to be a bridge between descriptive and inferential statistics (Webster, 1992). Without CLT, estimating population parameters from sample statistics would be impossible (Brightman, 1986). However, based upon our teaching and consulting experience, literature review, as well as our own research, we found that many students and researchers have serious misunderstandings of this theorem.

To counter this problem, there is an increasingly popularity in using dynamic computer software for illustrating CLT (e.g. Dambolena, 1984, 1986; Thomas, 1984; Bradley, 1984; Gordon and Gordon, 1989; Kerley, 1990; Myers, 1990; Bradley, Mittag, 1992; Hemstreet, & Ziegenhagen, 1992; Lang, 1993; Packard, 1993; Marasinghe, Meeker, Cook & Shin, 1994; Snyder, 1994). Packard (1993) found that learners were enthusiastic about using computer animation for learning CLT. Nonetheless, graphical displays do not necessarily clear up misconceptions related to this theorem (Myers, 1990).

In this paper, we will break down the theorem into several components, point out the common misconceptions in each part, and evaluate the appropriateness of use of computer simulation in the context of instructional strategies. The position of this paper is that even with the aid of computer simulations, instructors should explicitly explain the correct and incorrect concepts in each component of CLT.

Definition and Components

CLT states that a sampling distribution, which is the distribution of the means of random samples drawn from a population, becomes closer to normality as the sample size increases, regardless of the shape of distribution. As the name implies, CLT is <u>central</u> to large sample statistical inference and is true by <u>limitation</u>--it is true given that the sampling distribution is infinite. More generally, CLT tells that the distribution of a sum or average of a large number of independent random variables is close to normal. Based upon this definition, CLT can be divided into the following concepts and sub-concepts:

a) Randomness and random sampling

i. equality vs. independence

ii. non-self-correcting process

b) Relationships among sample, population and sampling distribution.

c) Normality

i. normal distribution as probability distribution (area = 100%)

ii. asymptotic tails

iii. symmetry and identical central tendencies

iv. two inflection points

d) Parameters of sampling distribution:

i. sample size

ii. population distribution

e) Relationships between sampling distribution and hypothesis testing

CLT is built upon the preceding concepts in a correlated manner. Misconceptions in one is likely to cause problems in others. For example, if one misconceives that random sampling is a self-correcting process, one may believe that the sampling distribution will compensate biases and magically restore equilibrium though the sample size is small or the distribution is non-normal. This oversimplified notion may lead one to overlook the asymptotic feature of normality--sampling distribution results from infinite numbers of sampling, as well as the mechanism of sampling distribution construction--sampling distribution is a function of both sample size and population. As a result, all these misconceptions together give one a false sense of security in conducting hypothesis testing with small and non-normal samples. Each of the above categories is addressed now:

Randomness and random sampling

Equality of chances. Many people define random sampling as a sampling process that each element within a set has equal chances to be drawn (e.g. Loether & McTavish, 1988; Myers, 1990; Moore & McCabe, 1993; Aczel, 1995). Equality is associated with fairness. This definition contributes to the myth that if the occurrence of a particular event is very frequent, the outcome is considered "unfair" and thus the sampling may not be random. This definition fails to reflect the reality that complete fairness virtually does not exist. For example, if a psycho-killer do a random shooting in a public area, children who have smaller bodies do not have equal chances to be shot as taller adults. By the same token, one should not expect that in an urn of

balls, small balls have equal probabilities to be sampled as large balls. Even if we put the same size balls in the urn, we cannot "equalize" all other factors that are relevant to the outcome. Jaynes (1995) fully explained this problem:

> The probability of drawing any particular ball now depends on details such as the exact size and shape of the urn, the size of balls, the exact way in which the first one was tossed back in, the elastic properties of balls and urn, the coefficients of friction between balls and between ball and urn, the exact way you reach in to draw the second ball, etc.. (Randomization) is deliberating throwing away relevant information when it becomes too complicated for us to handle...For some, declaring a problem to be 'randomized' is an incantation with the same purpose and effect as those uttered by an exorcist to drive out evil spirits...The danger here is particularly great because mathematicians generally regard these limit theorems as the most important and sophisticated fruits of probability theory. (pp. 319-320)

Before Jaynes, Poincare (1998) also made a similar criticism in a even more radical tone: "Chance is only the measure of our ignorance." (p.1359) Phenomena appear to occur according to equal chances, but indeed in those incidents there are many hidden biases and thus observers assume that chance alone would decide. Since authentic equality of opportunities and fairness of outcomes are not properties of randomness, a proper definition of random sampling should be a sampling process that each member within a set has <u>independent</u> chances to be drawn. In other words, the probability of one being sampled is not related to that of others.

At the early stage of the development of randomness, the essence of randomness was believed to be tied to independence rather than fair representation. It is important to note that when R. A. Fisher and his coworkers introduced randomization into experiment, their motive was not trying to obtain a representative sample. Instead they contended that the value of an experiment depends on the valid estimation of error (Cowles, 1989). In other words, the errors must be <u>independent</u> rather than systématic.

<u>Self-correcting process.</u> The ideas of equality and fairness in random sampling lead to another popular mind's bug: random sampling is a self-

correcting process i.e. different types of members of the population will eventually have a proportional representation in the sample (Tversky and Kahneman, 1982). Myers (1990) found that a computer simulation is not effective in removing the preceding misconception. Myers' finding is expected. First, a computer simulation tends to reinforce this erroneous belief. In a close system which is artificially generated by a computer, random sampling seems to restore equilibrium because variables in computer do not carry the complexity that are equivalent to the real world. Second, even a well-written software can mimic the perplexity of reality, "independence," "equality," "fairness," and "non-self-correcting" are too abstract to be illustrated in a computer environment. Carnap (1988) argued that one can generalize empirical laws, but not theoretical laws, from observing events and objects. It is unlikely that a learner will find out randomness does not guarantee equality and fairness after looking at the random sampling process on a computer screen many times.

Relationship between sample, population and sampling distributions

The concept of fair representation due to equal chances partly contributes to the confusion in the relationship among sample, population and sampling distribution. Many learners tend to perceive that a sample or a sampling distribution conforms to the shape of the population (Yu and Spencer, 1994). Further, students tended to forget the importance of random fluctuation and think of their sample as though it were the population. This leads to an under-appreciation of the role of probability in hypothesis testing as well, which will be discussed later. In addition, Yu and Behrens (1994) found that learners confused a sampling distribution with a sample i.e. learners failed to understand that a sampling distribution is constructed of the means from an infinite number of samples, instead of being a representation of a single sample. As stated earlier in the definition, CLT tells that the sum or average of a large number of independent random variables is approximately normally distributed. However, students failed to generalize beyond the sampling distribution of the mean (Yu & Behrens, 1994).

In order to remediate this misconception by computer simulation, the program should show the process of drawing samples from a huge population for constructing a sampling distribution. However, some programs show the outcome immediately after parameters have been entered, and leave the viewer to imagine the process (e.g. Thomas, 1984). Some provide only a small population

size and thus tends to encourage the confusion between sample and population (e.g. Mittag, 1992; Lang, 1993).

Normality

The confusion among sample, population and sampling distribution leads students to leave out important details of a normal sampling distribution such as asymptotic tails. There are four major properties in a normal curve. The feature that the total area underneath the curve is equal to 100 percent is straight-forward and causes less problems. Misunderstanding is likely to occur in asymptotic tails, symmetry and identical central tendencies, and inflection points.

Asymptotic tails. By definition, a sampling distribution is infinite and asymptotic. However, when showing a normal population, sample, and sampling distribution to students, most students were unable to distinguish the sampling distribution from the other two. When students were asked to draw a normal curve, most of them represented a normal curve as an inverted-U, failing to represent tails that extend without touching the x-axis (Yu & Spencer, 1994). This misconception may carry over to advanced statistical concepts. For example, Yu and Behrens (1994) found that quite a few students, including sophisticated learners, wondered why the normal curve in a power simulation does not touch the x-axis while power analysis is based on sampling distributions

Symmetry, identical central tendencies and inflection points. Most students know that in a normal curve, the two halves below and above the central tendencies are symmetrical, and the mean, mode, and median are all the same. However, this is a necessary but not sufficient condition of normality. Indeed, playtokurtic and leptokurtic distributions also share this characteristic. Because of this over-generalization, playtokurtic and leptokurtic distributions are sometimes misidentified as normal curves. As mentioned before, students usually represent a normal curve as an inverted-U (playtokurtic distribution). Many times both playtokurtic and leptokurtic distributions only have only one inflection point while a normal curve has two inflection points.

We do not have empirical evidence to decide whether visualization by computer could provide help in fixing these misconceptions. Many computer simulations do not show the property of asymptotic tails (e.g. Lang, 1993, Marasinghe et al. 1994). Some programs used histograms rather than smoothed

curves and therefore inflection points are not obvious. To remediate this problem, Wolfe (1991) overlaid a density curve on a histogram. However, even if a computer program does illustrate asymptotic tails and inflection points, we doubt if viewers can discover these subtle features.

As mentioned before, some programs use histograms instead of density curves to display distributions (e.g. kerley, 1990; Wolfe, 1991). Kerley (1990) acknowledged that the appearance of a histogram is tied to the number of intervals (bandwidth). It is quite often not apparent whether the distribution is normal or not. To compensate this limitation, kerley graphically illustrated the result of Kolmogorov-Smirnov normality test by confidence intervals. However, it may add more complexity to the problem by explaining Kolmogorov-Smirnov test and confidence bands.

Parameters of sampling distribution

Not only students had difficulties in identifying characteristics of normality, they also did not fully understand the parameters that determine the normality of a sampling distribution, which are sample size and population. Yu and Behrens (1994) found that quite a few students viewed a normal sampling distribution as the outcome of a magical sample size--30. This misconception is popularized by some statisticians who tried to set a cut-off point of large sample size for generating a bell-shaped sampling distribution out of a non-normal population. Some argued that the sample size should be at least 30, others have said 50 or 100, and some as high as 250 (cf. Saddler, 1971; Marasinghe et al. 1994). Gordon and Gordon (1989) argued that if the sample size is less than 30 and the population is not normal, the sampling distribution will be a t-distribution, otherwise, a normal distribution. Some authors who applied computer-based instruction to remediate misconceptions of CLT also erroneously adopted N=30 as the cut-off (e.g. Dambolena, 1986; Myers, 1990). The importance of the shape of the population is obscured with this kind of logic. We agree with Marasinghe et al. (1994) that there isn't any definite cut-off due to the infinite degrees of non-normality the underlying population could have. The worse the distortion from normality, the higher the sample size needed to form a normal sampling distribution.

A computer simulation is supposed to empirically show the above fallacy to learners. However, some programs provide only one population (e.g. Thomas, 1984; Lang, 1993). In some other programs neither the population nor the

sample size can be controlled by users (e.g. Dambolena, 1984). Consequently, they offer little help in clearing up this misunderstanding. By using a more versatile program *Statutor* (Wolfe, 1991), Yu and Spencer (1994) asked the subjects to manipulate different combinations of sample size and population. Students could immediately understand that the shape of the sampling distribution is a function of both parameters.

Relationship between sampling distribution and hypothesis testing

Hypothesis testing as probabilistic inference. All the previous misunderstandings partly lead to the improper application of CLT to hypothesis testing. When the relationship among population, sample and sampling distribution is not clear to students, most students failed to understand the rationale of hypothesis testing in terms of CLT. In hypothesis testing we usually draw only one sample out of the population and make inferences about the population based upon the single sample. Some students perceived that this sample alone can reflect the characteristic of the population without knowing that hypothesis testing is a probabilistic inference based upon the theoretical infinite number of samples with replacement.

It is important to note that CLT is based upon the notion that probability is a long run relative frequency, the ratio of the number of ways an event can occur to the total number of possible events (Cowles, 1989). People accept this "foundation" of hypothesis testing without realizing that there are other views to probability. Although James Bernoulli, one of the contributors to the development of CLT, accepted the inductive theory of probability, he was also attracted by the subjective concept of probability. Being bombarded by several incompatible ideas, he was not sure where to rest his case (cited in Cowles, 1989; Hacking, 1971). Skipping the history of CLT and probability in learning the relationships between the theorem and hypothesis testing may create an over secure sense of hypothesis testing.

False security of small and non-normal samples. The problem is perplexed by some statistical instructors and researchers who believe that CLT can "take care of" small or non-normal samples. For example, in the ListServe group *Stat-L*, a statistician said that a small sample size doesn't matter in hypothesis testing, because the *t* table and *F* table, which were computed based upon sampling distributions, will adjust the degrees of freedom for small samples. Moreover, a statistician at ACT stated that a non-normal sample will not affect parametric hypothesis testing, because according to CLT, the

sampling distribution will eventually become normal. We suspect that this misconception is related to the notion that "random sampling is a self-correcting process."

This represents a common confusion between the importance of the sampling distribution in hypothesis testing and the assumptions of a specific statistic (such as a t-test or an ANOVA). This is illustrated in a textbook: "Since the sampling distribution, and not the population, is used to test hypotheses, this means that whenever the sample is large we can completely relax the assumption about normality of the population and still use the known characteristics of the normal curve to test hypotheses." (cited in Brewer, 1985).

None of the software packages we reviewed addressed the above issues. This conceptual problem is more likely to be effectively taught by other media such as a lecture or a book rather than a computer package.

Conclusion

There is a strong support for the efficacy of dynamic computer graphic as an instructional medium (e.g. Flake, McClintock, & Turner, 1985; Riskin, 1990; Myers, 1990). The supporting theories that Myers (1991) adopted are Piaget's cognitive model and Constructivism. The former states that learners are benefited from linking abstract ideas with concrete representations. The latter endorses interactive media because students can construct knowledge through interaction rather than accepting information without deep processing.

There are some drawbacks in the above applications. By dividing CLT into components, it is obvious that some aspects of CLT can be clearly illustrated by computer software but some cannot. For example, dynamic computer graphic can perform the function of showing the process of sampling distribution. But it is difficult to present abstract concepts such as "equality," "independence," and relationship between CLT and hypothesis testing by graphical examples.

The constructivist approach gives learners much freedom to explore and construct ideas on their own. We are afraid that employing this approach to learn CLT may cause more misconceptions. As mentioned before, some subtle features such as asymptotic tails and two inflection points of a normal distribution cannot be easily observed by learners without explicit verbal instruction.

BEST COPY AVAILABLE

In conclusion, we recommend that in addition to computer-assisted instruction, statistical teachers should explicitly address all crucial aspects of CLT by telling both correct and incorrect concepts. Attention should be paid to the relationship between CLT and hypothesis testing because competent statisticians still commit mistakes in this regard.

References

Aczel, A. D. (1995). Statistics: Concepts and applications. Chicago: Richard D. Irwin, Inc.

Bradley, D. R. (1984 April). Computer simulation of classic studies in psychology. Paper presented at the Annual Meeting of the Eastern Psychological Association, Baltimore, MD.

Bradley, D. R., Hemstreet, R. L., & Ziegenhagen, S. T. (1992). A simulation laboratory for statistics. Behavior Research Methods, Instruments, & Computers, 24, 190-204.

Brewer, J. K. (1987). Behavioral statistics textbooks: Source of myths and misconceptions? Journal of Educational Statistics, 3, 252-268.

Brightman, H. J. (1986). Statistics in plain English. Cincinnati, OH: South-Western Publishing Co.

Carnap, R. (1988). The nature of theories. In Klemke, R. Hollinger, A. D. Kline (Eds.), Introductory readings in the philosophy of science (pp. 162-177). Buffalo, N.Y.: Prometheus Books.

Cowles, M. (1989). Statistics in psychology: An historical perspective. Hillsdale, New Jersey: LEA.

Dambolena, I. G. (1984). Teaching the central limit theorem through computer simulation. Mathematics and Computer Education, 18, 128-132.

Dambolena, I. G. (1986). Using simulation in statistical courses. Collegiate Computer, 4, 339-244.

Flake, J. L., McClintock, C. E., & Turner, S. V. (1985). Fundamentals of computer education. Belmont, CA: Wadsworth Publishing Company.

Gordon, F. S., & Gordon, S. P. (1989). Computer graphics simulations of sampling distributions. Mathematics and Computer Education, 21, 48-55.

Hacking, I. (1971). Jacques Bernoulli's art of conjecturing. British Journal Philosophy of Science, 22, 209-229.

Jaynes, E. T. (1995). Probability theory: The logic of science. [On-line]. Available World Wide Web: Hostname: http://www.math.albany.edu:8008/JaynesBook.html.

Kerley, L. M. (1990). Using computer graphics in statistics. Mathematics and Computer Education, 24, 232.

Lang, J. (1993). StatBox (HyperCard program). Valencia Community College.

Loether, H. J., & McTavish, H. J. (1988). Descriptive and inferential statistics: An introduction. Boston, MA: Allyn and Bacon, Inc.

Marasinghe, M., meeker, W., Cook, D., & Shin, T. S. (1994 August). Using graphics and simulation to teach statistical concepts. Paper presented at the Annual meeting of the American Statistician Association, Toronto, Canada.

Mittag, K. C. (192 April). Using computers to teach the concepts of the Central Limit Theorem. Paper Presented at the Annual Conference of the American Educational Research Association, San Francisco, CA (ERIC Document reproduction Service No. 349 947).

Moore, D. S. & McCabe, G. P. (1993). Introduction to the practice of statistics. New York: W. H. Freeman and Company.

Myers, K. N. (1990). An exploratory study of the effectiveness of computer graphics and simulations in a computer-student interactive environment in illustrating random sampling and the central limit theorem. Unpublished doctoral dissertation, Florida State University.

Packard, A. L. (1993 October). A comparison of three methods of teaching statistics. Paper presented at the Annual Conference of the American Educational Research Association, Chicago, IL (ERIC Document reproduction Service No. 365 696).

Poincare, H. (1988). Chance. In J. Newman (Ed.), The world of mathematics (Vol. 2). (pp. 1359-72). Redman, WA: Tempus.

Riskin, S. R. (1990). Teaching through interactive multi-media programming: A new philosophy of the social sciences and a new epistemology of creativity. California. (ERIC Document reproduction Service No. 327 133).

Saddler, D. R. (1971 Sept.). The central limit theorem--An empirical investigation. The Australian Mathematics Teacher, p.90-94.

Snyder, R. (1994 November). Applications of a computer Algebra system in introductory statistics. Paper presented at the Annual meeting of the Society for Computing in Psychology, St Louis, MO.

Thomas, D. A. (1984). Understanding the central limit theorem. Mathematical Teacher, 77, 542-543.

Tversky, T. & Kahneman, D. (1982). Judgment under certainty: Heuristics and Biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under

uncertainty: heuristics and biases (pp. 3-19). Cambridge: Cambridge Univ. press.

Yu, C. H. & Behrens, J. (1994 August). Misconceptions in statistical Power and dynamic graphics as a remediation. Poster session presented at the Annual Meeting of American Statistical Association, Toronto, Ontario, Canada.

Yu, C. H., & Spencer, A. (1994). An instructional design system of using computer simulations for teaching the Central Limit Theorem. Unpublished manuscript.

Webster, E. (1992). Evaluation of computer software for teaching statistics. Journal of computers in mathematics and science teaching, 11, 377-391.

Wolfe, R. A. (1991). Statutor (Computer program). Department of Biostatistics, School of Public Health, University of Michigan.